The Journal of Social Sciences Research
ISSN(e): 2411-9458, ISSN(p): 2413-6670
Vol. 6, Issue. 4, pp: 483-493, 2020
URL: https://arpgweb.com/journal/journal/7
DOI: https://doi.org/10.32861/jssr.64.483.493

Academic Research Publishing Group

Original Research Open Access

Permutation of UTME Multiple-Choice Test Items on Performance in Use of English and Mathematics Among Prospective Higher Education Students

Bassey A. Bassey
Department of Educational Foundations, University of Calabar, Calabar, Nigeria

Isaac O. Ubi
Department of Educational Foundations, University of Calabar, Calabar, Nigeria

German E. Anagbougu
Department of Educational Foundations, University of Calabar, Calabar, Nigeria

Valentine J. Owan*
Department of Educational Management, University of Calabar, Calabar, Nigeria

Abstract

In an attempt to curtail examination malpractice, the Joint Admission and Matriculation Board (JAMB) has been generating different paper types with a different order of test items in the Unified Tertiary Matriculation Examination (UTME). However, the permutation of test items may compromise students' performance unintentionally because constructive suggestions in theory and practice recommend that test items be sequenced in ascending order of difficulty. This study used data collected from a random sample of 1,226 SSIII students to ascertain whether the permutation of test items has any effect on the performance of students in two different subjects (Use of English and Mathematics). The study adopted the Equivalent Groups Quasi-Experimental Research Design with three independent groups. Findings emerged, amongst others, that there is a significant difference in the performance scores of prospective university students' in use of English and Mathematics examintaions arranged in three different orders (ED, DE, R). There are no significant gender differences in the performance of students in Use of English and Mathematics based on test item permutation. However female students perform better than male students when test items are arranged in ascending order of difficulty while males perform better when test items are arranged in descending order of difficulty. It was concluded that the permutation of test items in UTME examination tends to affect the performance of students in Use of English and Mathematics. This finding has implications for the future conduct of UTME examinations and enrolment into higher education as the randomization of UTME test items changes the difficulty order of different paper types. It was recommended that other measures of curtailing examination malpractices that would not affect students' academic performance should be adopted.

Keywords: Permutation; UTME; Test-items; Performance; Multiple-choice; English; Mathematics.

1. Introduction

The use of tests in educational or behavioural disciplines is to assess the academic performance or learning outcomes of individuals and to assess curriculum and educational needs of an academic programme (Çokluk *et al.*, 2016). Thus, decisions can be made from the results of tests for educational advancement or improvement. Test results form a critical decision base to judge the capabilities of different individuals in terms of their knowledge, understanding and skills. To Çokluk *et al.* (2016), test results aid in understanding individuals, employing, placing, selecting, guiding and assessing them. Little wonder why many educational institutions are using this vital tool for grading students' performance. Consequently, it is very necessary for educators to develop test and prove empirically (through psychometric approaches) that such tests, as may have been designed, possess high validity and reliability. Some scholars warned that individual/organizational test developers, practitioners, and interpreters must develop and implement eligible methods to examine test development and psychometric qualifications (Camilli and Shepard, 1994; Holland and Wainer, 1993).

Soureshjani (2011), posits that test construction includes two major aspect - the "trait" (the knowledge intended to be measured) and the "method" (the approach through which the trait is measured''. The trait can be measured using different methods which can lead to different results in the performance of test-takers. It is well documented and widely known that the performance of students on a test is affected by numerous factors (Soureshjani, 2011). These numerous factors have been widely studied and considered in the literature in both theoretical and practical studies (Owan *et al.*, 2019; Owan *et al.*, 2020a; Robert and Owan, 2019; Soureshjani, 2011). This study is aimed at evaluating the methods used by the JAMB in testing prospective higher education students during placement examination.

The Unified Tertiary Matriculation Examination (UTME) is an annual entrance examination conducted, regulated and monitored by the Joint Admissions and Matriculations Board (JAMB) for prospective undergraduates

*Corresponding Author

into Nigerian tertiary institutions. These tertiary institutions include private and public universities, polytechnics and colleges of education. Candidates qualified to take this examination are those who have obtained certificates from either one or all of the West African Examinations Council (WAEC) and the National Examination Council (NECO). The minimum standard for admission into different categories of tertiary institutions are presented hereunder: Public Universities (Minimum of 160 marks); Private Universities (Minimum of 140 marks); Public Polytechnics (Minimum of 120 marks); Private Polytechnics (Minimum of 110 marks); and Colleges of Education (Minimum of 100 marks) (JAMB, 2020). However, institutions have the liberty to raise their cut-off marks for various courses above the minimum standard set by JAMB.

The test items in this examination are usually multiple-choice items. The examination mode was formerly Paper-Pencil Test (PPT) until in 2013 when it was changed to Computer-Based Test (CBT) (Nnaike and Ogundare, 2016). Three modes (Dual-Based Test (DBT), PPT and CBT) were used in 2013 and 2014; while in 2015, the board dwelt solely on CBT (Nnaike and Ogundare, 2016), which has continued to the year 2020. The switch from PPT to CBT was initiated by the JAMB as a strategy to curtail issues of examination malpractice, time consumption in the administration of the test, and high demand for human and material resources which usually follows PPT (Adebayo, 2016). Another reason for the switch has also been suspected to be ease in scoring of test items, which is more efficient with the use of computers than manually. So far, the use of CBT has been widely commended due to the effective deployment of ICT facilities for the examination, provision of safe and relaxed environments that is free from external influences, disturbances and interference by anyone (Adebayo, 2016).

Apart from the measures outlined above, JAMB introduced the "paper type" approach as another means of addressing the issue of examination malpractice. The paper type measure allows JAMB to scramble or arrange test items in different orders. Thus, the first item in a student script may be the fiftieth item on the script of another. Different paper types come with different arrangements of test items, such that two candidates sitting close to each other do not have the opportunity of copying from one another. However, it seems to the researchers that such a practice by JAMB to curtail examination malpractice is good, but fear that the approach may have an implication on the students' performance in the test. Çokluk *et al.* (2016), disclosed that changing the order of test items potentially leads to adverse outcomes for examinees, although the primary reason for such an application is to prevent cheating in examinations. A general principle in measurement and evaluation is to start a test with easy items before difficult ones (i.e. to arrange them in ascending order of difficulty). Contravening this principle might increase test anxiety and poor learning outcomes. The same test arranged in different item orders may technically be seen as different tests.

Taking the same test with items permutated in different orders may also pose different levels of anxiety on students. A study documents that an easy-to-difficult (ED) test leads to lower anxiety levels than a difficult-to-easy (DE) test (Çokluk *et al.*, 2016). The argument here is that some studies have documented that test items arranged in different orders affect the performance of students, completion time and items parameters (Balch, 1989; Carlson and Ostrosky, 1992; Çokluk *et al.*, 2016; Hauck *et al.*, 2017; Jessel and Sullin, 1975; Owan *et al.*, 2020a; Picou and Milhomme, 1997; Soureshjani, 2011; Towle and Merrill, 1975). Jessel and Sullin (1975), also found that there is a significant difference in the performance and reliability of tests based on test items arrangement. Towle and Merrill (1975), found that students scored significantly higher in tests whose items are arranged from easy to hard (EH) than arranged from hard to easy (HE). The study further documents that test-item difficulty sequence is not correlated with anxiety arousal (Towle and Merrill, 1975).

However, other studies (such as Gohman and Spector (1989); Carlson and Ostrosky (1992) disagreed partly with the results of Jessel and Sullin (1975) by showing evidence that students' test scores may be influenced by the sequence of the items but not the validity and reliability of the items. Other studies (Çokluk *et al.*, 2016; Newman *et al.*, 2009; Owan *et al.*, 2019; Soureshjani, 2011) provided further evidence that there is a disparity in the performance of students taking easy to difficult (ED) multiple-choice test and those taking difficult to easy (DE) items.

The results in these studies indicated that students who took ED items outperformed those who took DE items. From these studies cited, it suggests that the performance of students in UTME may be affected as candidates are usually given randomly-arranged questions. Therefore, the randomization and permutation of UTME test items may lead to biased measurement output which can be favourable to some candidates and unfavourable to others. The reason is that different items orders are known to have different variance forms, which could cause students to consider items as "difficult" or "easy" (Balch, 1989; Impara and Foster, 2006; Laffittee, 1984; Newman *et al.*, 2009; Pettijohn and Sacco, 2007). Although previous studies have focused more on test item arrangement based on the order of difficulty (ED and DE), little focus has been paid to randomly-arranged test items.

In terms of the interaction between gender and test item permutation, some studies document that the interaction between gender and test item permutation does not significantly influence students' performance in Mathematics (Owan *et al.*, 2020b; Schee, 2012). In a related study, Lane *et al.* (1987) reported that there is a significant effect of gender on performance of students based on cognitive difficulty labels (knowledge, comprehension, and application). Bielinski and Davison (1998), conducted two studies to test for gender-by-item-difficulty interaction on performance in Mathematics using data from nine forms of a basic skills test in mathematics. It was discovered that males tended to outperform females on the hardest items; females tended to outperform males on the easiest items (Bielinski and Davison, 1998). Plake *et al.* (1983), used performance feedback for a nonquantitative examination based on a Latin Square design and provided evidence that significant sex-by-order effects did not occur. The lack of this effect was attributed to the use of a nonquantitative examination (Plake *et al.*, 1983). The authors gave a call for

the use of quantitative examination to ascertain if differential effects of item feedback may be a source of explanation for evidence of sex-by-order effects (Plake *et al*., 1983).

The present study investigated the interactive effects of gender and test item permutation using quantitative techniques based on this call. Previous studies have focused so much on the effects of test item arrangement on performance in Mathematics with little or no focus on performance in use of English. To be specific, the permutation of UTME test items and its effect on performance has never been assessed in the Nigerian and foreign literature, as previous studies had use other national qualifying examinations or self-made test developed by the researchers. UTME is a national qualifying examination in the Nigerian context that students take to gain admission into higher education. Therefore, a study of this magnitude is highly necessary to enable JAMB conduct a fair examination to all concerned. Furthermore, no study has focused on the performance of prospective students which has implications on future enrolments into higher education. These are some key gaps the present study has been designed to fill while contributing to the arguments currently ongoing in the literature. Thus, the specific questions this study has been designed to answer are:

i. What is the mean performance of prospective university students in the use of English examination arranged in three different orders (ED, DE, R)?
ii. What is the mean performance of prospective university students in Mathematics examination arranged in three different orders (ED, DE, R)?
iii. What is the mean performance of male and female prospective higher education students in Use of English based on test-items permutation?
iv. What extent is the mean performance of male and female prospective higher education students in Mathematics based on test-items permutation?

1.1. Statement of Hypotheses

The following hypotheses were formulated and tested in this study.

i. There is no significant difference in the performance scores of prospective university students' in the use of English examination arranged in three different orders (ED, DE, R).
ii. There is no significant difference in the performance scores of prospective university students' in Mathematics examination arranged in three different orders (ED, DE, R).
iii. There is no significant effect of gender and test-items permutation on the performance of prospective higher education students in the Use of English.
iv. Gender and test item permutation does not significantly influence prospective higher education students in mathematics.

2. Methods

2.1. Research Design

This study adopts the Equivalent Groups Quasi-Experimental Research Design, with three independent groups. In using this design, the researchers randomly assigned the sampled respondents into three independent groups that were exposed to three examination conditions (based on item order). The equivalent groups' approach was used because the process of grouping respondents was based on randomization. The random grouping of individuals into various groups was to ensure that participants of similar attributes are present across groups.

2.2. Population Sampling Procedures Sample

This study targeted a population of 9,768 SSIII (Senior Secondary Class Three) students distributed across 88 public secondary schools in Calabar Education Zone. The population of SSIII was considered in the study because they are in the final year of secondary education and are likely to take UTME after completion. Thus, they are referred to, in this study, as prospective higher education students because upon passing of the UTME, many of them will be enrolled into higher education institutions. Three-stage clustered sampling technique was adopted in selecting the study's sample. The first stage involves the clustering of schools based on the Local Education Authorities (LEAs) in the education zone. There are seven LEAs in Calabar Education Zone out of which four LEAs were randomly selected. In the second stage, 50% of all the schools available in each selected LEA were randomly selected. In the third and last stage, all the SSIII students in each of the selected schools were included as participants of the study. Consequently, a sample of 1,226 SSIII students in 26 public secondary schools was studied as shown in Table 1.

Table-1. Sample distribution of the study showing the number of schools and SSIII students studied

LEAs	Number of schools available	50% of available schools selected	Population of SS3 students
Calabar Municipality	16	8	353
Akamkpa	19	10	491
Akpabuyo	7	4	174
Calabar South	8	4	208
Total	50	26	1226

Source: Cross River State Secondary Education Board (2019)

A randomized approach was used in each school to split the entire population of SSIII students into three equivalent groups. In achieving this, the attendance register was used to obtain the list of SSIII students. With the list, three groups were formed (A, B, C) by assigning the first, second and third student on the list to the first, second and third group respectively. Similarly, the fourth, fifth and sixth student were assigned to group A, B and C respectively. This pattern continued until all the students were assigned to the various groups available. The groups were mutually exclusive as no student was assigned to more than one group. In general, 410 students were assigned to Group A (ED), 409 were assigned to Group B (DE), and 407 were assigned to Group C (R) as shown in Table 2.

Table-2. Distribution showing the classification of the study's participants into various groups

LEAs	GROUPS			Total
	A (ED)	B (DE)	C (R)	
Calabar Municipality	118	118	117	353
Akamkpa	164	164	163	491
Akpabuyo	58	58	58	174
Calabar South	70	69	69	208
Total	410	409	407	1226
ED = Easy-to-Difficult; DE = Difficult-to-Easy; R = Randomised Items				

Source: Authors' field survey (2019)

2.3. Research Instrument

The instrument used for data collection was UTME past question in Mathematics and Use of English for the year 2018. This instrument was adopted since the respondents are SS3 students and have never sat for any UTME examinations before. The instrument was also used since the items therein have been used by JAMB in 2018. The items in the past question for both Mathematics and Use of English were re-arranged in three orders – ED, DE, and R. Each of the three different test items orders served as a paper type. Thus, there were three paper types ED, DE, and R, with each paper type having three sections. Section A of each paper type elicited respondents' demographic data (Gender and Age), section B and C of each paper type were questions pertaining to Mathematics (40 items) and Use of English (60 items), uniquely arranged in an order based on paper type.

The instruments were not validated by the researchers since JAMB is a criterion-referenced examination with psychometric experts who already validated the items. The reliability of these various paper types was, however, ascertained using Test-Retest technique. In achieving this, a trial test was conducted using 60 SS3 students in Calabar Education (These were part of the population but not the sample). These respondents were chosen since they were part of the population and shared similar characteristics as those in the sample. The 60 respondents were split into three random groups of 20 participants each. The first, second and third group responded to the ED, DE and R items respectively. After an interval of two weeks, the same respondents were given the same paper types to answer. At the end of the exercise scores obtained from both sets of administration were correlated using Pearson Product Moment Correlation. The result of the reliability analysis is presented in Table 3.

Table-3. Test-retest reliability results of the study

Subject	Paper type	Administration	N	K	$\bar{X}$	SD	r
Mathematics	ED	First	20	40	24.60	7.022	.865
		Second	20	40	24.65	6.434	
	DE	First	20	40	16.70	7.012	.802
		Second	20	40	18.30	7.364	
	R	First	20	40	13.40	7.170	.871
		Second	20	40	15.10	7.490	
Use of English	ED	First	20	60	34.55	11.709	.818
		Second	20	60	34.30	10.016	
	DE	First	20	60	25.55	13.036	.897
		Second	20	60	29.20	14.111	
	R	First	20	60	22.95	9.795	.770
		Second	20	60	24.30	8.105	
N = No of respondent; K = No of items; R = Reliability coefficient; $\bar{X}$ = Mean; SD = Standard deviation							

Source: Field Survey (2019)

2.4. Procedure for Data Collection/Analysis

Data were collected from the participant at the various Local Education Authorities (LEAs) selected for the study on different days. At each LEA, the three independent groups of SS3 students were identified and made to sit in a hall. Corresponding paper types (ED, DE, and R) were distributed to each group of participants. The sitting arrangement was randomly done (students were made to sit between groups) typical of UTME. The test lasted for an hour and 30 minutes after which the administered question papers were retrieved. The entire data collection exercise took two weeks to complete since the researchers visited two LEAs per week. Due to the rigorous approach adopted, it was possible for the researchers to retrieve all the administered questions (representing 100 per cent rate of return.

Descriptive statistics such as mean and standard deviation were used to analyse data, as well as, answer the research questions. The hypotheses of this study were tested at the .05 alpha level using inferential statistics such as one-way analysis of variance (ANOVA) and Analysis of Covariance (ANCOVA) where applicable. All computations were aided using SPSS version 25 software.

3. Results

The results of this study are presented in line with the research questions and hypotheses guiding the study.

3.1. Research Question

What is the mean performance of prospective university students in the use of English examination arranged in three different orders (ED, DE, R)? The answer to this research question was provided using the results presented in Table 4.

Table-4. Mean performance and standard deviation of prospective university students in the use of English examination arranged in three different orders

Test item permutation	N	$\bar{X}$	SD	SE	Min.	Max.
ED	410	34.24	15.292	.755	3	60
DE	409	28.65	15.346	.759	1	60
R	407	27.78	15.830	.785	1	60
Total	1226	30.23	15.741	.450	1	60

The result contained in Table 4 indicates that 410, 409, and 407 SS3 students respectively, responded to ED, DE, and R arranged test items in the use of English. Students who took the easy-to-difficult (ED) permutation achieved a mean score of 34.24 and standard deviation of 15.292. Those who took the difficult-to-easy test items (DE) had a mean score of 28.65 and standard deviation of 15.346. Students who took the randomly (R) arranged test items achieved a mean score of 27.78 and standard deviation of 15.830. This result suggests that students' performance in the use of English is higher where test items were arranged in ascending order of difficulty. This is followed by a descending order of difficulty and random permutation. When we use the group mean score of 30.23 as a criterion, the result in Table 4 shows that only the mean score for ED beats the criterion mean, while the means for DE and R are well below the criterion mean. This suggests that students' performance in ED is relatively high while the performance in DE and R respectively, is relatively low. Table 4 also showed that the minimum score recorded by a student in ED category is 3 marks, while the score recorded in DE and R categories is 1 mark respectively. The maximum score recorded across the three test permutations was 60 marks.

3.2. Research Question Two

What is the mean performance of prospective university students in Mathematics examination arranged in three different orders (ED, DE, R)? This research question was answered using the result of descriptive statistics presented in Table 5.

Table-5. Mean performance and standard deviation of prospective university students in Mathematics examination arranged in three different orders

Test item permutation	N	$\bar{X}$	SD	SE	Min.	Max.
ED	410	25.31	9.344	.461	3	40
DE	409	22.73	10.704	.529	3	40
R	407	18.56	10.035	.497	1	40
Total	1226	22.21	10.413	.297	1	40

As shown in Table 5, students who sat for the ED arranged test items in Mathematics achieved a mean of 25.31 and a standard deviation of 9.344. Those who sat for the DE and R arranged test items in Mathematics achieved means of 22.73 and 18.56 with a standard deviation of 10.704 and 10.035 respectively. The minimum score recorded in Mathematics for ED and DE class is 3 marks respectively, while the minimum score recorded for R is 1 mark. However, the maximum score recorded across the three categories is 40 marks. Furthermore, the result shown in Table 5 indicates that students' performance in Mathematics is higher where test items were arranged in ED order. This is followed by performance where test items are arranged in DE and R in that order. Using the overall group mean of 22.21 as a criterion, it can be inferred from the results in Table 5 that students' Mathematics performance in ED and DE permutation is relatively high, while the performance in R category is relatively low.

3.3. Research Question Three

What is the mean performance of male and female prospective higher education students in the Use of English based on test-items permutation? Descriptive statistics such as mean and standard deviation were used to provide an answer to this research question. This was achieved through a crosstab of gender versus test item permutation shown in Table 6.

Table-6. Mean performance and standard deviation of male and female prospective higher education students in Use of English based on test-items permutation

Gender	Test item permutation	$\bar{X}$	SD	N
Male	ED	33.50	16.124	185
	DE	28.65	15.141	199
	R	27.99	15.861	202
	Total	29.95	15.865	586
Female	ED	34.85	14.582	225
	DE	28.65	15.574	210
	R	27.58	15.835	205
	Total	30.49	15.634	640
Total	ED	34.24	15.292	410
	DE	28.65	15.346	409
	R	27.78	15.830	407
	Total	30.23	15.741	1226

The result in Table 6 shows that a total of 586 male took the use of English test. Out of these, 185 were in the ED category, 199 were in the DE category, and 202 were in the R category. For females, a total of 640 prospective higher education students sat for the use of English test. Out of these, 225 were in the ED category, 210 were in the DE category, and 205 were in the R category. Generally, male prospective higher education students achieved a mean of 29.95 in use of English with a standard deviation of 15.865, while female prospective university students achieved an overall mean of 30.49 in use of English with a standard deviation of 15.634. This result implies that female prospective higher education students outperformed their male counterparts in the use of English.

Specifically, the result in Table 6 shows that male students scored a mean of 33.50 and a standard deviation of 16.124 in the ED permutation in the use of English. Female students scored a mean of 34.85 and a standard deviation of 14.582 in the ED permutation in the use of English. Thus, in the ED category, female students performed better than male students. In the DE permutation, the result in Table 6 shows that male students achieved a mean of 28.65 and standard deviation of 15.141, while female students also had a mean of 28.65 and a standard deviation of 15.574. This result shows evidence that there is no difference in the performance of male and female prospective university students in DE permutation in the use of English. In the R permutation, male prospective students achieved a mean score of 27.99 with a standard deviation of 15.861 while female prospective university students achieved a mean score of 27.58 with a standard deviation of 15.835. Comparatively, it can be inferred that students' performance in the use of English based on R arrangement is higher for males than females.

3.4. Research Question Four

What extent is the mean performance of male and female prospective higher education students in mathematics based on test-items permutation? The answer to this research question was provided using the mean and standard deviation scores of male and female students based on the gender versus test items permutation crosstab in Table 7.

Table-7. Mean performance of male and female prospective higher education students in mathematics based on test-items permutation

Gender	Test item permutation	$\bar{X}$	SD	N
Male	ED	26.04	9.214	185
	DE	23.21	10.636	199
	R	18.09	9.935	202
	Total	22.34	10.475	586
Female	ED	24.71	9.427	225
	DE	22.28	10.774	210
	R	19.01	10.136	205
	Total	22.09	10.363	640
Total	ED	25.31	9.344	410
	DE	22.73	10.704	409
	R	18.56	10.035	407
	Total	22.21	10.413	1226

The results presented in Table 7 shows that out of the 586 male prospective higher education students selected for this study, 185 of them took the ED Mathematics test items, 199 took the DE Mathematics items, and 202 took the R Mathematics test items. Also, a total of 640 female prospective higher education students sat for the Mathematics test. Out of these, 225 were in the ED category, 210 were in the DE category, and 205 were in the R category. Generally, Male prospective higher education students achieved a mean and standard deviation of 22.34 and 10.475 in the Mathematics test while female prospective higher education students scored a mean and standard deviation of 22.09 and 10.363 respectively in the Mathematics test. This result implies that the performance of male prospective higher education students is better than that of their female counterparts in Mathematics.

Specifically, the result in Table 7 showed that males achieved a mean of 26.04 with a standard deviation of 9.214 in the ED category, while female scored a mean of 24.71 with a standard deviation of 9.427 in the ED

category in Mathematics. This result indicates that male students performed better than females based on ED permutation in Mathematics. For DE permutation, male prospective higher education students had a mean performance of 23.21 with a standard deviation of 10.636, while their female counterpart recorded a mean of 22.28 with a standard deviation of 10.774 in Mathematics. Thus, males outperformed female prospective higher education students in the DE category in Mathematics. For R permutation, a mean score and standard deviation of 18.09 and 9.935 was recorded for male prospective higher education students, while a mean score and standard deviation of 19.01 and 10.136 was recorded for their female counterparts in Mathematics. Therefore, the performance of females is higher than males in the R category.

3.5. Hypothesis One

There is no significant difference in the performance scores of prospective university students' in the use of English examination arranged in three different orders (ED, DE, R). This null hypothesis was tested at the .05 alpha level using one-way analysis of variance. The result of the analysis is presented in Table 8.

Table-8. Summary of one-way analysis variance result of the difference in the performance of prospective university students in the use of English based on test item permutation

Source of variation	SS	Df	MS	F	Sig.
Between Groups	10047.399	2	5023.700	20.936	.000
Within Groups	293468.813	1223	239.958		
Total	303516.212	1225			

As presented in Table 8, the result revealed an F-ratio of 20.936 with a p-value of .000 which is below the alpha level of .05 at 2 and 1223 degrees of freedom. Based on this result, the null hypothesis was rejected while the alternate hypothesis is upheld. This implies that there is a significant difference in the performance scores of prospective university students' in the use of English examination arranged in three different orders (ED, DE, R). Thus, the mean difference recorded in Table 4 was not due to chance. However, in ascertaining which of the mean differences of the various pair-wise comparison was statistically significant, the Tukey Honest Significant Difference (HSD) test of multiple comparisons was performed as shown in Table 9.

Table-9. Tukey HSD of multiple comparisons of the three different test permutation (ED, DE, R) in the use of English

(I) Test item permutation	(J) Test item permutation	$\bar{X}$ Difference (I-J)	SE	Sig.
ED	DE	5.586*	1.083	.000
	R	6.458*	1.084	.000
DE	R	.871	1.085	.701

*. The mean difference is significant at the 0.05 level

The Tukey HSD multiple comparisons presented in Table 9 shows that there is a significant mean difference (5.586, $p<.05$) in the performance of prospective university students who took the ED and those who took the DE in use of English. When students who took ED were compared with those who took R, the result in Table 9 revealed that students who took ED significantly outperformed their R counterparts, with a mean difference of 6.458. Lastly, the comparison between the performance of students in the DE and R categories revealed that there is a mean difference of .871 which is not statistically significant. By conclusion, the use of English performance of prospective university students in the ED permutation is significantly higher than that of the DE and R permutation respectively. While there is no significant difference in the performance in the use of English between prospective higher education students in DE and R permutation.

3.6. Hypothesis Two

There is no significant difference in the performance scores of prospective university students' in Mathematics examination arranged in three different orders (ED, DE, R). One-way analysis of variance was employed in testing this hypothesis at the .05 level of significance. The result of the analysis is shown in Table 10.

Table-10. One-way analysis variance result of the difference in the performance of prospective university students in Mathematics based on test item permutation

Source of variation	SS	Df	MS	F	Sig.
Between Groups	9493.001	2	4746.500	47.064	.000
Within Groups	123340.961	1223	100.851		
Total	132833.962	1225			

The result in Table 10, disclosed that there is an F-ratio of 47.064 with a p-value of .000 which is less than the .05 level of significance at 2 and 1223 degrees of freedom. Consequently, the null hypothesis was rejected giving sufficient statistical evidence to uphold the alternate hypothesis. By implication, there is a significant difference in the performance scores of prospective university students' in Mathematics examination arranged in three different orders (ED, DE, R). Hence, the observed mean differences recorded in Table 5 was not by chance. The Tukey HSD test of pair-wise comparison was performed to show which of the mean differences were significant (see Table 11).

Table-11. Tukey HSD of multiple comparisons of the three different test permutation (ED, DE, R) in Mathematics

(I) Test item permutation	(J) Test item permutation	$\bar{X}$ Difference (I-J)	SE	Sig.
ED	DE	2.581*	.702	.001
	R	6.757*	.703	.000
DE	R	4.176*	.703	.000

*. The mean difference is significant at the 0.05 level

From Table 11, it was discovered that the Mathematics performance of prospective higher education students in the ED category is significantly higher than that of students in the DE (mean difference = 2.581, p<.05) and R (mean difference = 6.757, p<.05) test item permutation categories respectively. Furthermore, the result showed that there is a significant mean difference (4.176, p<.05) in the Mathematics performance of prospective higher education students in the DE and R categories.

3.7. Hypothesis Three

There is no significant effect of gender and test-items permutation on the performance of prospective higher education students in the Use of English. This hypothesis was tested using the result of the test between-subject effect of the Analysis of Covariance presented in Table 12.

Table-12. Summary of between-subject effects of gender and test item permutation on prospective higher education students' performance in Use of English

Source	Type III SS	Df	MS	F	Sig.
Corrected Model	10249.525[b]	5	2049.905	8.528	.000
Intercept	1114518.877	1	1114518.877	4636.439	.000
Gender	30.375	1	30.375	.126	.722
Test item permutation	9766.650	2	4883.325	20.315	.000
Gender * Test item permutation	171.980	2	85.990	.358	.699
Error	293266.687	1220	240.383		
Total	1424022.000	1226			
Corrected Total	303516.212	1225			
b. R Squared = .034 (Adjusted R Squared = .030)					

The results of the analysis presented in Table 12 shows that gender has no significant main effect on prospective university students in Use of English {F (1, 1220) = 0.126, p>05}. This implies that the mean differences recorded for male and female prospective students in Table 6 are not statistically significant. Test item permutation has a significant main effect on prospective higher education students' performance in Use of English {F (2, 1220) = 20.315, p<.05}. This result justifies the ANOVA result in Table 8. The result in Table 12 shows a non-significant effect of the interaction between gender and test item permutation on the performance of prospective higher education students in Use of English {F (2, 1220) = 0.358, p>.05}. Thus, there is no significant difference in the performance scores of male and female prospective higher education students in Use of English based on test items permutation. Based on this result, the null hypothesis was retained implying that there is no significant effect of gender and test-items permutation on the performance of prospective higher education students in the Use of English. However, both gender and test item permutation explained 3.4% of the total variance in the performance of prospective higher education students in Use of English with the remaining 96.6% of the total variance explained by other independent variables not included in the model.

3.8. Hypothesis Four

Gender and test item permutation does not significantly influence prospective higher education students in mathematics. The analysis of covariance results presented in Table 13 was used in testing the hypothesis at the .05 level of significance.

Table-13. Analysis of Covariance results of the interaction between gender and item permutation on performance in Mathematics

Source	Type III SS	df	MS	F	Sig.
Corrected Model	9847.757[a]	5	1969.551	19.538	.000
Intercept	603450.006	1	603450.006	5986.110	.000
Gender	60.185	1	60.185	.597	.440
Test item permutation	9635.835	2	4817.918	47.793	.000
Gender * Test item permutation	294.559	2	147.280	1.461	.232
Error	122986.205	1220	100.808		
Total	737491.000	1226			
Corrected Total	132833.962	1225			
a. R Squared = .074 (Adjusted R Squared = .070)					

The result of ANCOVA depicted in Table 13 shows that main effect of gender on the performance of prospective higher education students in Mathematics {F (1, 1220) = 0.597, p>.05). This implies that male did not differ significantly from females in their performance in Mathematics. The main effect of test item permutation on the performance in mathematics of prospective higher education students is statistically significant {F (2, 1220) = 47.793, p<.05}. This implies that there are significant differences in the performance in Mathematics of prospective higher education students in the ED, DE and R categories (see Table 11). The interactive effects of gender and test item permutation on the performance of prospective higher education students in Mathematics is not statistically significant {F (2, 1220) = 1.461, p>.05}. Consequently, the null hypothesis was retained implying that gender and test item permutation does not significantly influence prospective higher education students in mathematics. However, gender and test item permutation contributed 7.4% to the total variance in the performance of prospective higher education students in Mathematics with the remaining 92.6% of the variance attributed to other independent variables not included in the model.

4. Discussion

The first major finding of this study established that there is a significant difference in the performance scores of prospective university students' in use of English examination arranged in three different orders (ED, DE, R). It was also uncovered that the use of English performance of prospective university students in the ED permutation is significantly higher than that of the DE and R permutation respectively. While there is no significant difference in the performance of prospective higher education students in DE and R permutation in the use of English between. This finding is not surprising since the best practice in the principle of measurement recommends that test items be arranged in ascending order of difficulty. This finding aligns with the result of Towle and Merrill (1975) that students scored significantly higher in tests whose items are arranged from easy to hard (EH) than arranged from hard to easy (HE). The differences in the performance scores of students in the use of English in the DE and R categories may be attributed to the permutation of test items. DE test items may increase the level of anxiety in students resulting in poor outcomes in the use of English. Çokluk *et al.* (2016) discovered that an easy-to-difficult (ED) test leads to lower anxiety levels than a difficult-to-easy (DE) test. Thus, anxiety is increased when students take the DE or R arranged items as opposed to the ED permutation.

The second main finding of this study showed that there is a significant difference in the performance scores of prospective university students' in Mathematics examination arranged in three different orders (ED, DE, R). It was also discovered that the Mathematics performance of prospective higher education students in the ED category is significantly higher than that of students in the DE and R test item permutation categories respectively. Furthermore, there is a significant mean difference in the Mathematics performance of prospective higher education students in the DE and R categories. Just as in the first finding, the effect of test anxiety when students face DE and R arranged items may be the reason for the differences in the performance of students in Mathematics. This finding corroborates the result of previous studies such as Soureshjani (2011); Çokluk *et al.* (2016); Hauck *et al.* (2017); Owan *et al.* (2020b), which document that test items arranged in different order affects the performance of students.

The third main finding of this study shows that there is no significant effect of gender and test-items permutation on the performance of prospective higher education students in the Use of English. The study showed that female prospective higher education students outperformed their male counterparts in the use of English. This finding agrees with the results of Lane *et al.* (1987), however, the difference in the performance scores of male and female students based on test items permutation in use of English is not statistically significant. The insignificance of the mean difference in the performance of male and female students implies that the permutation of test items affects both gender at the same rate. In the ED category, female students performed better than male students. There is no difference in the performance of male and female prospective university students in DE permutation in the use of English. In the R permutation, students' performance in the use of English is higher for males than females. Female may have performed insignificantly better than males in the use of English because many female students tend to make English language their preferred subject and perhaps due to their good accent in pronunciation. Their accent in pronouncing words coupled with their interest in English may improve their motivation to study more, resulting in good academic performance. Another reason may be the attraction of many females towards art courses that requires good performance in English language to thrive such as literature, mass communication, theatre art, law, history, English and literary studies and so on. Male may have done better in the randomly arranged test in the use of English because the arrangement of test item at random transposes the difficulty order. Thus, many females may become bored when faced with difficult items first, while males may appear to pay a greater level of attention in such a situation. This aligns with the result of Bielinski and Davison (1998) which discovered that males tended to outperform females on the hardest items; females tended to outperform males on the easiest items (Bielinski and Davison, 1998).

The fourth finding of this study showed that gender and test item permutation does not significantly influence the performance of prospective higher education students in mathematics. Although the performance of male prospective higher education students is better than that of their female counterparts in Mathematics, the mean difference is not significant. This tallies with the results of previous studies (Owan *et al.*, 2020b; Schee, 2012), suggesting that permutating test items in mathematics examination affects the performance of both male and female students at the same rate. Furthermore, it was revealed that male students performed better than females based on the ED and DE permutation in Mathematics while females outperformed males in the R permutation. Males have performed insignificantly better than females perhaps because many male students tend to be attracted to science-related disciplines such as computer science, engineering, Mathematics, Statistics, and so on. Such attraction to the

sciences may stir up their interest and motivation in learning mathematics giving them an edge over their female counterparts. Due to the paucity of research literature on gender and test item permutation, the finding of this study has left a challenge for prospective researches to be conducted to validate the findings documented in this study.

5. Conclusion

Based on the findings of this study, it was concluded that the permutation of test items in the Unified Tertiary Matriculation Examination (UTME) tends to affect the performance of students in the Use of English and Mathematics. There are no significant gender differences in the performance of students in Use of English and Mathematics based on test item permutation. However female students perform better than male students when test items are arranged in ascending order of difficulty while males perform better than female students when test items are arranged in descending order of difficulty in the Use of English. Males perform better than female students when test items are arranged in both ascending and descending order of difficulty in Mathematics. This finding has implication for the future conduct of UTME examinations and enrolment into higher education as the randomization of UTME test items changes the difficulty order and reliability of different paper types. This leads to differences in performance and consequently, the enrolment prospect into higher education. Teachers, as well as test and measurement experts, also need to be aware of the role gender plays in the ordering of test items. Based on this conclusion, it was recommended that:

i. The Joint Admission and Matriculation Board (JAMB) should stop the permutation of test items (resulting in different paper types) as a means of preventing examination malpractice. This would help in the improvement of prospective higher education students' performance in Use of English and Mathematics, as well as their enrolment into higher education;
ii. Other measures of curtailing examination malpractices that would not affect students' academic performance in key or minor subjects should be adopted. Such measures may include the use of CCTV cameras, mixing science and art students together in the same examination halls, providing adequate invigilators, and punishing students caught cheating.
iii. Teachers and measurement experts should ensure that test item are always arranged in ascending order of difficulty. In gender-sensitive cases, test items should be arranged from difficult-to-easy for girls and from difficult-to-easy for boys.

References

Adebayo, A. (2016). JAMB CBT: Why nigeria must move forward with technology. Available: https://www.thecable.ng/jamb-cbt-nigeria-must-move-forward-technology

Balch, W. R. (1989). Item order affects performance on multiple-choice exams. *Teaching of Psychology,* 16(2): 75–77. Available: https://doi.org/10.1207/s15328023top1602_9

Bielinski, J. and Davison, M. L. (1998). Gender differences by item difficulty interactions in multiple-choice Mathematics items. *American Educational Research Journal,* 35(3): 455–76. Available: https://doi.org/10.3102/00028312035003455

Camilli, G. and Shepard, L. A. (1994). *Methods for identifying biased test items.* Sage: London, UK.

Carlson, J. L. and Ostrosky, A. L. (1992). Item sequence and student performance on multiple-choice exams: Further evidence. *The Journal of Economic Education,* 23(3): 232-35. Available: https://doi.org/10.1080/00220485.1992.10844757

Çokluk, Ö., Gül, E. and Doğan-Gül, Ç. (2016). Examining differential item functions of different item ordered test forms according to item difficulty levels. *Educational Sciences: Theory and Practice,* 16: 319-30. Available: http://dx.doi.org/10.12738/estp.2016.1.0329

Gohman, S. F. and Spector, L. C. (1989). Test scrambling and student performance. *The Journal of Economic Education,* 20(3): 235–38. Available: https://doi.org/10.2307/1182298

Hauck, K. B., Mingo, M. A. and Williams, R. L. (2017). A review of relationships between item sequence and performance on multiple-choice exams. *Scholarship of Teaching and Learning in Psychology,* 3(1): 58–75. Available: https://doi.org/10.1037/stl0000077

Holland, P. W. and Wainer, H. (1993). *Differential item functioning.* Lawrence Erlbaum Associates Publishers: New Jersey, NJ.

Impara, J. and Foster, D., 2006. "Strategies to minimize test fraud." In *S. M. Downing & T. M. Haladyna (Eds.), Handbook of test development. Mahwah, NJ: Lawrence Erlbaum Associates.* pp. 91–114.

Jessel, J. C. and Sullin, W. L. (1975). The effect of Keyed Response Sequencing of multiple-choice items on performance and reliability. *Journal of Educational Measurement,* 12(1): 45-48. Available: https://doi.org/10.1111/j.1745-3984.1975.tb01008.x

Laffittee, R. G. (1984). Effects of item order on achievement test scores and students' perceptions of test hardy. *Teaching of Psychology,* 77(4): 212–14.

Lane, D. S., Bull, K. S., Kundert, D. K. and Newman, D. L. (1987). The effects of knowledge of item arrangement, gender, and statistical and cognitive item difficulty on test performance. *Educational and Psychological Measurement,* 47(4): 865–79. Available: https://doi.org/10.1177/0013164487474002

Newman, D. L., Kundert, D. K., Lane, D. S. and Bull, K. S. (2009). Effect of varying item order on multiple-choice test scores: Importance of statistical and cognitive difficulty. *Applied Measurement in Education,* 1(1): 89-97. Available: https://doi.org/10.1207/s15324818ame01018

Nnaike, U. and Ogundare, F. (2016). Nigeria: Jamb utme-stakeholders prefer cbt despite hitches. Available: https://allafrica.com/stories/201603311144.html

Owan, V. J., Etudor-Eyo, E. and Esuong, U. U. (2019). Administration of punishment, students' test anxiety, and performance in Mathematics in secondary schools of Cross River State, Nigeria. *International Journal of Academic Research in Business and Social Sciences,* 9(6): 415–30. Available: https://doi.org/10.6007/IJARBSS/v9-i6/5963

Owan, V. J., Bassey, B. A. and Ini, S. E. (2020a). Interactive effect of gender, test anxiety, and test items sequencing on academic performance of ss3 students in mathematics in calabar education zone, Cross River State, Nigeria. *American Journal of Creative Education,* 3(1): 21–31. Available: https://doi.org/10.20448/815.31.21.31

Owan, V. J., Bassey, B. A. and Agurokpon, D. C. (2020b). Path analytic study of factors affecting students' attitude towards test-taking in secondary schools in afikpo education zone, Ebonyi State, Nigeria. *American Journal of Creative Education,* 3(1): 10–20. Available: https://doi.org/10.20448/815.31.10.20

Pettijohn, T. F. and Sacco, M. F. (2007). Multiple-choice exam question order influences on student performance, completion time, and perceptions. *Journal of Instructional Psychology,* 34(3): 142–49.

Picou, A. and Milhomme, A. J. (1997). The effect of random and sequential versions on student test performance. *Financial Practice and Education,* 7: 85–90.

Plake, B. S., Melican, G. J., Carter, L. and Shaughnessy, M. (1983). Differential performance of males and females on easy to hard item arrangements: Influence of feedback at the item level. *Educational and Psychological Measurement,* 43(4): 1067–75. Available: https://doi.org/10.1177/001316448304300416

Robert, I. A. and Owan, V. J. (2019). Students' perception of teachers' effectiveness and learning outcomes in Mathematics and Economics in secondary schools of Cross River State, Nigeria. *International Journal of Contemporary Social Science Education,* 2(1): 157–65.

Schee, B. A. V. (2012). Test item order, level of difficulty, and student performance in marketing education. *Journal of Education for Business,* 88(1): 36–42. Available: https://doi.org/10.1080/08832323.2011.633581

Soureshjani, H. K. (2011). Item sequence on test performance: Easy items first? *Language Testing in Asia,* 1(3): 46–59. Available: https://doi.org/10.1186/2229-0443-1-3-46

Towle, N. J. and Merrill, P. F. (1975). Effects of anxiety type and item-difficulty sequencing on Mathematics test performance. *Journal of Educational Measurement,* 12(4): 241-49. Available: https://doi.org/10.1111/j.1745-3984.1975.tb01025.x